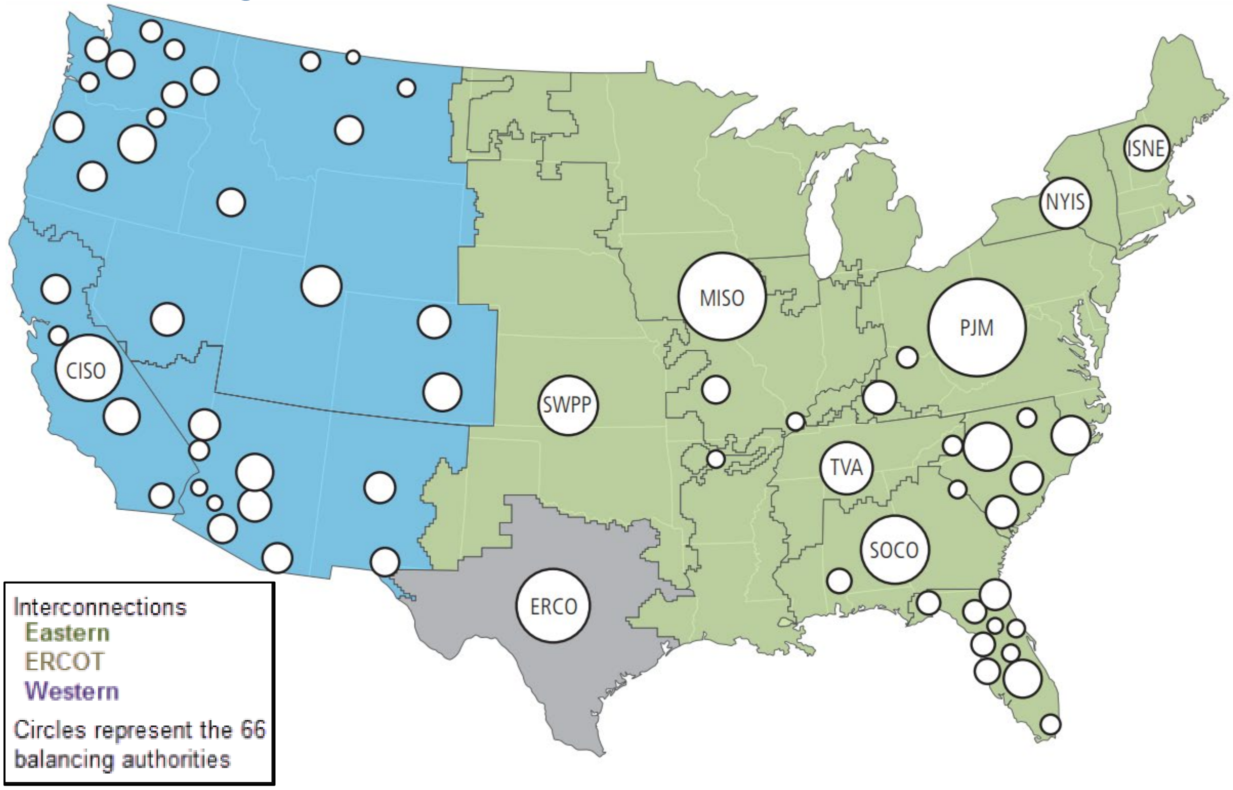
How it Works: The Role of a Balancing Authority

There are many key players in the electricity supply chain that work together to reliably supply power to electric customers. One key player is the balancing authority, which manages the operation of the electric system within a specific geographic area. There are more than 60 balancing authorities in the U.S., and they are typically either utilities, Power Marketing Administrations (PMAs), or a group of utilities that have formed regional entities called regional transmission organizations (RTOs) and independent system operators (ISOs).[1]

Exhibit 1. Balancing Authorities Across the U.S.



Source: Quadrennial Energy Review: Transforming the Nation's Electricity System: The Second Installment of the QER (January 2017)

See EIA Hourly Electric Grid Monitor for interactive map.

A balancing authority ensures that power system demand and supply are always balanced, which maintains safe and reliable operation of the power system. If supply falls below demand, it can cause a drop in frequency below the stable system frequency, risking significant permanent physical damage to electrical equipment and prolonged grid-wide blackouts. During such situations, balancing authorities may direct utilities to "shed load" (i.e., implement rolling blackouts) to bring supply and demand into balance.

[1] U.S. Energy Information Administration Form 930

Balancing authorities also manage transfers of electricity (interchanges) with other balancing authorities and use economic dispatch to optimize the use of various generating units to minimize real-time costs. Balancing authorities are responsible for maintaining operating conditions under mandatory reliability standards issued by the North American Electric Reliability Corporation (NERC) and approved by the U.S. Federal Energy Regulatory Commission (FERC).

ISOs/RTOs

The formation of ISOs and RTOs came at the encouragement of FERC (click to read on FERC orders and RTOs/ISOs) to operate the transmission system independently of—and foster competition for electricity generation among—wholesale market participants. ISOs/RTOs do not own transmission or generation assets, nor do they directly serve or establish pricing for retail electric customers. They are not a part of any government, so they also do not have any enforcement authority over any electric generation facilities, electric transmission or distribution lines, or substations.[2]

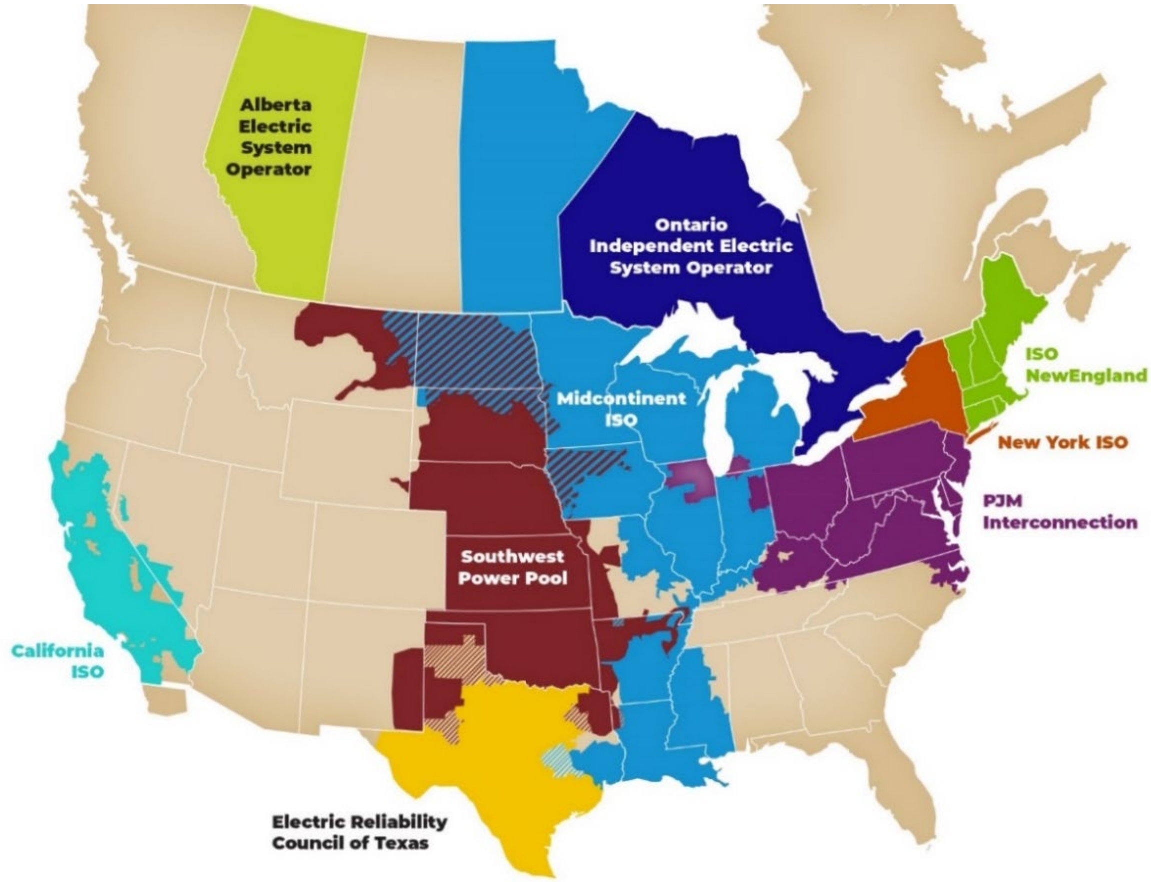
The role of ISOs and RTOs are similar, but ISOs do not meet the minimum requirements specified by FERC to hold the designation of RTO or have not applied for the status. An ISO operates the region's electricity grid, oversees the region's wholesale electricity markets, and provides reliability planning for the region's bulk electricity system. RTOs perform the same functions as the ISOs but have greater responsibility for planning near- and long-term transmission maintenance, upgrades, and expansion projects in collaboration with the actual transmission system owners. Utilities in areas where there is no RTO or ISO serve all these functions for their own territories. There are currently three ISOs and four RTOs that operate within the U.S.:

Independent System Operators (ISO)	California ISO	CAISO
	New York ISO	NYISO
	Electric Reliability Council of Texas[3]	ERCOT
Regional Transmission Operator (RTO)	PJM Interconnection	PJM
	Midcontinent Independent System Operator	MISO
	Southwest Power Pool	SPP
	ISO New England	ISO-NE

Although two-thirds of the nation’s electricity load is served in RTO/ISO regions (as can be seen from the map in Exhibit 2), there are large sections of the United States—particularly in the Southeast and the West—where there is no ISO or RTO. Electric utilities in these areas are still subject to the same reliability regulations under NERC/FERC.

[2] The North American Electric Reliability Corporation (NERC) issues and enforces reliability standards for balancing authorities, which are approved by the U.S. FERC and, in Canada, by Canadian regulators. NERC relies on Regional Entities to enforce reliability standards with bulk power system owners.
[3] ERCOT does not fall under FERC authorities over interstate transmission and wholesale markets since it is entirely within the state of Texas and is not interconnected with the rest of the region. ERCOT is still subject to NERC oversight and FERC regulation for reliability.

Exhibit 2. Map of ISO/RTO Territories



Source: ISO/RTO Council

Grid Operation

Balancing authorities function as grid operators that dispatch electric generators to provide reliable power at the lowest cost. Because each generator has differing variable costs, generation is dispatched using the least costly generator first, in a way that is consistent with the relevant constraints of the transmission system and reliability requirements. Traditional utilities that are balancing authorities manage economic dispatch within their service areas, while ISOs and RTOs determine economic dispatch using bid-based markets where buyers and sellers bid for or offer generation. The dispatch process occurs in two phases: day-ahead unit commitment for the next day's dispatch, and real-time system dispatch.

- **Day-Ahead Commitment:** Balancing authorities decide which generating units should be committed to be online for each hour of the next 24-hour period. In selecting the most economic generators to commit, operators consider forecasted load requirements and each unit's physical operating characteristics, such as how quickly output can be changed, maximum and minimum output levels, and the minimum time a generator must run once it is started. Load forecasting uses mathematical models to predict demand across the balancing authority's footprint, incorporating weather/climate, past energy consumption data, typical behavioral patterns, and economic activity.
- **Real-Time System Dispatch:** Balancing authorities must decide the appropriate level at which each available resource from the day-ahead commitment should be operated, given the real-

time load and other grid conditions, so that overall production costs are minimized. Actual demand will vary from the forecasted load and the balancing authorities must adjust the actual dispatch to match. As part of real-time operations, demand, generation, and interchange (imports and exports) must be continually kept in balance to maintain the stable system frequency. This is typically done by automatic generation control to change the generation dispatch as needed.

Contingency Resources and Capacity Markets

Grid operators maintain excess contingency resources in the form of power generation capacity, intertie[4] capacity, and electricity storage to ensure that electricity supply is always available to meet anticipated peak demand. This "reserve margin" is required to meet Federal and State reliability requirements and can be used to provide a cushion or buffer during unexpected demand spikes, or potential loss of supply or transmission resources. If reserve margins fall below specified levels, grid operators may initiate emergency procedures to use additional resources—such as ancillary services and load management measures—that are only available during scarcity conditions.

RTO/ISOs typically use capacity markets to ensure that there are sufficient resources available to serve anticipated load and meet reserve margin requirements. Capacity markets may use auctions to lock in prices for electric capacity from generation resources in advance of the need. Capacity markets can also be marketplaces for demand response, in which customers reduce demand when called upon in exchange for capacity payments similar to what generators receive. Capacity prices vary by location and timing of capacity commitments, and typically not by size or fuel type. ISO-NE, PJM, MISO, and NYISO operate capacity markets, while other ISOs currently do not.

Ancillary Services

Ancillary services are intended to maintain electric reliability and support electricity transmission. They mainly include services related to frequency stability like operating reserves; services related to voltage stability like reactive power support and voltage control, and other services like black start generation. Ancillary services are becoming more important for variable energy resources—such as wind and solar power—due to their production variability and forecast uncertainty. Within ISOs/RTOs, ancillary services are typically procured through bid-based auction markets, tariff fixed or formula-based rates, or competitive solicitations.

Operating reserves are needed to quickly restore load and generation balance when a supply resource trips offline. Operating reserves are provided by dispatchable generating units increasing output or demand-side resources reducing demand to restore balance. There are three types:

1. **Spinning reserves** are provided by generators that are online (synchronized to the system frequency) with some spare capacity and capable of increasing its electricity output within a short time period (within 10 minutes).
2. **Non-spinning reserves** are provided by generating units that are not necessarily synchronized to the power grid but can be brought online within a short time period (within 10 minutes). Non-spinning reserves can also be provided by demand-side resources.
3. **Supplemental reserves** are provided by generating units that can be made available within a longer time period (between 10 to 30 minutes).

[4] An intertie is an interconnection permitting passage of current between two or more electric utility systems

Black start generating units—primarily hydroelectric facilities and diesel generators—can transition from a shutdown condition to an operating condition and start delivering power without any outside assistance from the electric grid. These are the first facilities to start up in the event of a system failure or blackout to restore the rest of the grid.

Emergency Measures

Balancing electricity supply and demand to preserve grid stability is a critical function of balancing authorities. If supply falls below demand, it causes a frequency drop on the electrical grid below the stable system frequency, which can cause vibrations resulting in significant permanent physical damage to electrical equipment. As the frequency drops below safe operating levels, many generators will automatically disconnect from the grid for self-protection. As generators and other electrical equipment shut down, this further reduces electricity supply, causing the system frequency to drop even further.

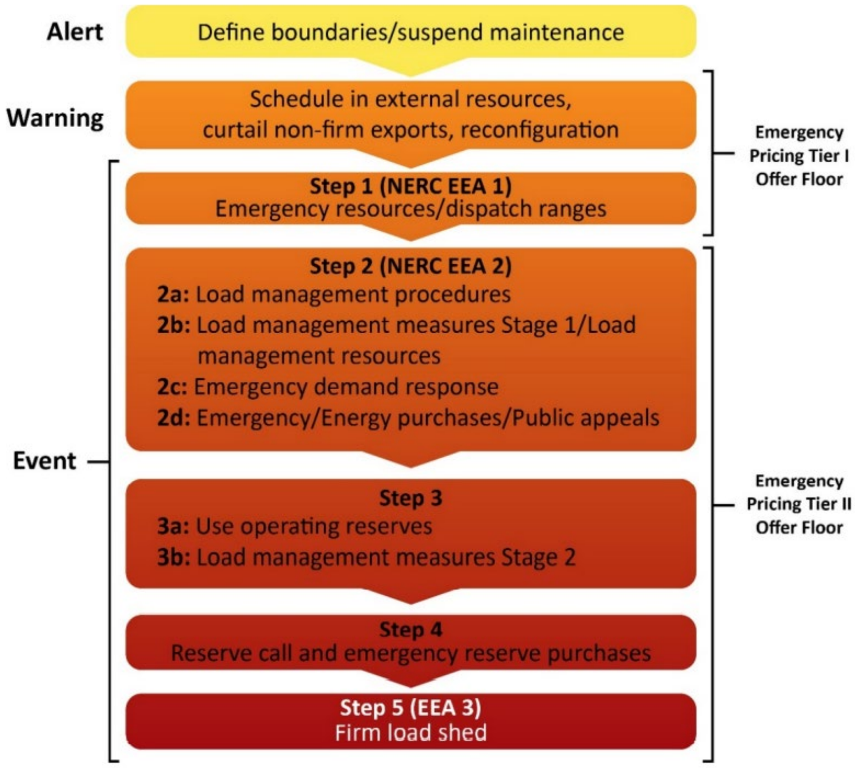
If no interventions are taken to stabilize the grid, this cascading effect can lead to collapse of the electricity grid and prolonged grid-wide blackout. During the February 2021 Extreme Cold & Winter Weather Incident, significant electricity generation assets went offline in the ERCOT region, causing a frequency drop that triggered additional generation loss. To rebalance supply and demand to stabilize the grid, ERCOT had to quickly shed substantial load, calling on its member utilities to proactively cut off electricity to their customers on a rotating basis over a period of three days. Although grid conditions in neighboring balancing areas were not as severe as in ERCOT, both MISO and SPP implemented brief periods of load shedding during the event.

When a balancing authority predicts there will be a shortage of electricity, there are response measures they can take to rebalance supply and demand. Each balancing authority has its own response protocol with response measures increasing as the anticipated shortage becomes more severe. The first step is typically to issue an alert to market participants describing the situation and suspending non-essential work (such as regular maintenance) to maximize the availability of generation and transmission resources. Balancing authorities may also issue Conservation Alerts, in which they ask consumers to voluntarily reduce their energy usage to reduce demand, especially during peak demand hours. Examples of consumers reducing energy usage include adjusting their thermostats, avoiding using large electric appliances, turning off non-essential lights, etc.

If initial response measures are insufficient to rebalance supply and demand, then a balancing authority will begin declaring energy emergency alerts (EEAs). EEAs have different levels depending on the event.

- At EEA Level 1, a balancing authority will typically call on all available power supplies (regardless of economics), including power from other grids if available, and enable certain demand-side resource deployments.
- At EEA Level 2, a balancing authority will typically initiate demand response programs that shed load from large industrial customers that have contractually agreed. They will also typically make a public appeal for all customers to cut back on electricity consumption.
- At EEA Level 3, balancing authorities will typically order utilities to shed load and inform them how much load they individually need to shed, and each utility then sheds the load by disconnecting circuits around customers with firm contracts. These outages are often "rotated" through the utility's service territory with customers losing power for several hours at a time, although sometimes customers are without power for the entire event. During load sheds, utilities typically will not disconnect circuits serving critical customers (hospitals, water treatment plants, etc.).

Exhibit 3. Emergency Alert Levels



Source: NERC